SPLITTING AND SUCCESSIVELY SOLVING AUGMENTED LAGRANGIAN METHOD FOR OPTIMIZATION WITH SEMICONTINUOUS VARIABLES AND CARDINALITY CONSTRAINT[*]

YANQIN BAI[†], RENLI LIANG[‡], AND ZHOUWANG YANG[§]

Abstract. We propose a new splitting and successively solving augmented Lagrangian (SSAL) method for solving an optimization problem with both semicontinuous variables and a cardinality constraint. This optimization problem arises in several contexts such as the portfolio selection problem, the compressed sensing problem and the unit commitment problem, etc. The problem is in general NP-hard. We derive an optimality condition for this optimization problem, under some suitable assumptions. By introducing an auxiliary variable and using an augmented Lagrangian function, the constraints are decomposed into two parts. By fixing particular variables, the optimization problem is split into two subproblems. The first one is an easy solving convex programming problem, while the second one is a complicated quadratic programming problem with semicontinuous variables and cardinality constraint. The two subproblems are solved alternatively. The second subproblem is transformed equivalently into a mixed integer quadratic programming problem. Based on its features of both the objective function and the constraints, we solve the mixed integer quadratic programming problem successively, then obtain its exact optimal solution directly. Furthermore, we prove the convergence of SSAL, under some suitable assumptions. Finally, we implement our method for the portfolio selection problem and the compressed sensing problem, respectively. Real world data and simulation study show that SSAL outperforms the well-known CPLEX 12.6 and the penalty decomposition (PD) method, while enjoying a similar cardinality of decision variable. For example, the numerical results are even nearly 200 times faster than that of CPLEX 12.6 for the portfolio selection problem, and more than 40 times faster than PD method for the compressed sensing problem, respectively. In particularly, SSAL is powerful when the size of problem increase largely.

Key words. semicontinuous variables and cardinality constraint optimization problem, alternating direction method of multipliers, augmented Lagrangian decomposition,

AMS subject classifications. 90C11, 90C27, 90C30

1. Introduction.

In this paper we consider the following minimization problem:

$$
\begin{array}{lll}
(\mathrm{P}) & \min & f(x) \\
 & \text{s.t.} & x \in X, \\
 & & \|x\|_0 \le K, \\
 & & x_i \in \{0\} \cup [a_i, b_i], \quad i \in \{1, 2, \ldots, n\},
\end{array}
$$

where $x \in R^n$, $f(x)$: $R^n \to R$ is a continuously differentiable convex function and $X \in R^n$ is a closed convex set with no empty interior, which represents the general constraints for x. $\|x\|_0$ denotes the cardinality (i.e., the number of nonzero entries) of the vector x. Furthermore $0 < K < n$, $0 < a_i < b_i$. A variable $x_i \in \{0\} \cup [a_i, b_i]$ is referred to as a semicontinuous variable.

The problem (P) arises in several contexts such as the portfolio selection problem (see [5], [13], [16]), the compressed sensing problem (see [4], [7]), the separable

[*] This research was supported by a grant from the National Natural Science Foundation of China (No.11371242)

[†] Department of Mathematics, Shanghai University, Shanghai,200444, China. (yqbai@shu.edu.cn)

[‡] Department of Mathematics, Shanghai University, Shanghai, 200444, China. (13720011@shu.edu.cn)

[§] School of Mathematical Sciences, University of Science and Technology of China, Hefei, 230026, China (yangzw@ustc.edu.cn)

quadratic facility location problem [24] and the unite commitment problem (see [15]), ect. Obviously, the constraints of (P) are both nonconvex and nonsmooth. Problem (P) is a highly nonlinear optimization problems and may have many local minimizers. So, it is difficult to solve problem (P), even if to find its local solutions. In general, problem (P) is NP hard [5]. For decades, several analytical and approximate methods, including reformulation, approximation and relaxation, have been developed for solving problem (P).

Padberg and Rinaldi [21] firstly proposed a branch-and-cut algorithm for the symmetric traveling salesman problem, which is an integer linear programming problem. Frangioni and Gentile [15] developed the branch-and-cut algorithm for a general program with a convex objective function and the semicontinuous constraint, by using a perspective function to derive lower bounds on the objective function value. Zheng, Sun and Li devolved the SDP approach by applying a special Lagrangian decomposition scheme for a quadratic programming problem with the cardinality and the semicontinuous constraints, which gives the tightest continuous relaxation of the perspective reformulation of the problem in [25]. The above exact methods are able to solve only a fraction of practically useful models, a variety of heuristic procedures have also been proposed. Chang et al. presented three heuristic algorithms based upon genetic algorithms, tabu search and simulated annealing for the standard mean variance portfolio optimisation model with cardinality constrained in [10]. There are also various approximate methods and techniques for dealing with the cardinality or sparse constraint. Burdakov et al presented a reformulation for the cardinality-constrained problems, including the semi-continuous case and they formulated the problems into a mixed-integer ones in literature [8]. Moreover, they also introduced a mixed-integer formulation whose standard relaxation still has the same solutions (in the sense of global minima) as the underlying cardinality-constrained problem [9]. Furthermore, they derived the suitable stationarity conditions and suggested an appropriate regularization method for the solution of optimization problems with cardinality constraints. Bai et al proposed an alternating direction method of multipliers (ADMM) method for $\ell_1 - \ell_2$ regularized logistic regression model in [2]. And Bai and Shen developed this method for the consensus proximal support vector machine for classification in [3].

Recently, Lu and Zhang [18] proposed a novel penalty decomposition (PD) method for an l_p norm minimization problem which minimizes a general nonlinear convex function subject to $\|x\|_0 \leq K$. Besides the cardinality constraints, Cui et al. [13] investigated a portfolio selection problem with both the cardinality constraints and the semicontinuous constraints, which was reformulated as a mixed integer quadratically constrained quadratic program (MIQCQP). They devolved the branch-and-bound method by reformulating MIQCQP to a new MIQCQP reformulation, which has a more tighter continuous relaxation bound than the MIQCQP. They reported the computational results for problems with up to 400 assets. Although the speed of the PD method is generally faster than the hard-thresholding algorithm [6], the solution obtained by the PD method is not better enough for the solution quality. Besides, the reformulation of Cui et al. [13] has more constraints and variables than MIQCQP and caused the difficulty that the branch-and-bound method is slow and fails to solve large-scale problems.

Inspired by requirements of both large-scale problems and the better solution quality, we propose a new method for solving problem (P), which covers the portfolio selection problem and the compressed sensing problem as special cases. Under mild assumptions,we derive the optimality conditions of problem (P). By introducing an auxiliary variable and using an augmented Lagrangian function, the constraints are separated into two parts. By fixing particular variables, the optimization problem is split into two subproblems. The first one is an easy to solve convex programming problem, while the second one is a complicated quadratic programming problem with semicontinuous variables and cardinality constraint. Therefore, the second subproblem is transformed into an equivalent mixed integer quadratic programming problem with $0-1$ knapsack constraint. Then we propose the augmented Lagrangian alternating direction method and solve the two subproblems alternatively. The second one can be solved analytically, due to features of both the objective function and the constraints. Furthermore, we prove that our method is convergent. Finally, we implement our method for the portfolio selection problem and the compressed sensing problem. The numerical results demonstrate that our method is faster than several existing methods. For example, in case of the compressed sensing problem, our method is more than 40 times faster than that of the penalty decomposition (PD) method in [18].

The paper is organized as follows. In Section 2, we describe some examples of problem (P) arising from different real-world applications. In Section 3, we establish the first order optimality conditions for the problem (P). In Section 4 we present our method for solving problem (P). We name it the Splitting and Successively solving Augmented Lagrangian method (abbreviated as SSAL). We also establish some convergence results. In Section 5, we conduct numerical experiments to test the performance of SSAL for the portfolio selection problem in [13] and compressed sensing problems. Finally, we present some conclusions in Section 6.

1.1. Notation. In this paper, the symbols R^n and R^n_+ denote the n-dimensional Euclidean space and the nonnegative orthant of R^n, respectively. For any vectors $x = [x_1, x_2, \dots, x_n]^T$, $y = [y_1, y_2, \dots, y_n]^T$, $x \cdot y = [x_1y_1, x_2y_2, \dots, x_ny_n]^T$, and $x^2 = x \cdot x$. Given an index set $L \subseteq \{1, 2, \dots, n\}$, $|L|$ denotes the size of L, $\bar{L}$ is the complement of L in $\{1, 2, \dots, n\}$. x_L denotes the subvector formed by the entries of x indexed by L. Likewise, X_L denotes the submatrix formed by the columns of X indexed by L. We denote by I the identity matrix, whose dimension should be clear from the context. Given a closed set $C \subseteq R^n$, $N_C(x)$ and $T_C(x)$ denote the normal and tangent cone of C at any $x \in C$, respectively. For any real vector, $\|\cdot\|_0$ and $\|\cdot\|$ denote the cardinality (i.e., the number of nonzero entries) and the Euclidean norm of the vector, respectively.

2. Examples of Applications.

In this section, we describe some examples of problem (P) arising from the portfolio selection, the compressed sensing and the subset selection. More applications can be found in [24].

2.1. Portfolio selection problem.

Let μ and Q be the mean vector and the covariance matrix of n assets with return $r = (r_1, r_2 \cdots r_n)^T$, respectively. Let $x = (x_1, x_2 \cdots x_n)^T$ be the portfolio weights to the n risky assets. In real-world, most investors only choose a small number of stocks to invest. In other words, we need to consider the constraint $\|x\|_0 \le K$ to control the total number of different assets in the optimal portfolio. We also need to prevent the investors from holding some assets with very small amount due to the transaction cost and managerial concerns. So we need to consider the semicontinuous constraints: $x_i \in \{0\} \cup [a_i, b_i], i \in \{1, 2, \dots, n\}$. A mean-variance portfolio selection model with semicontinuous and cardinality constraints can be modeled as:

$$\begin{aligned} \text{(SP)} \quad & \min \quad x^T Q x \\ & \text{s.t.} \quad \mu^T x \geq \rho_0, \\ & \qquad e^T x = 1, \\ & \qquad \|x\|_0 \leq K, \\ & \qquad x_i \in \{0\} \cup [a_i, b_i], \quad i \in \{1, 2, \ldots, n\}. \end{aligned}$$

where $e = (1, 1 \cdots 1)^T$, ρ_0 is the prescribed return level.

2.2. Compressed sensing problem. Compressed sensing (CS) is an important problem in signal processing (see, for example, [11]).The CS problem with nonnegativity constraints can be formulated as

$$\begin{aligned} \text{(NCS)} \quad & \min \quad \frac{1}{2}\|Ax - b\|^2 \\ & \text{s.t.} \quad \|x\|_0 \leq K, \quad x \geq 0, \end{aligned}$$

where $A \in R^{p \times n}$ is a data matrix, $b \in R^p$ is an observation vector, K is an integer for controlling the sparsity of the solution. More applications and reformulations of problem (NCS) can be found in [20] and [17]. In the compressed sensing problem, it is often assumed that $p < n$.

In multivariate linear regression, we are given p observed data points (a_i, b_i), with $a_i \in R^n$ and $b_i \in R$. The goal is to minimize the least square measure of $\sum_{i=1}^p (a_i^T x - b_i)^2$ with only a subset of the prediction variables in x. This subset selection problem then has the same form as the problem (NCS). In contrast with the case of compressed sensing, the number of data in subset selection is often much larger than the dimension of the data ($p > n$).

In practice, we can always impose an upper bound on x, i.e. $x \leq u$, for some sufficiently large positive numbers u (see [24]).

$$\begin{aligned} \text{(NCSB)} \quad & \min \quad \frac{1}{2}\|Ax - b\|^2 \\ & \text{s.t.} \quad \|x\|_0 \leq K, \quad 0 \leq x \leq u. \end{aligned}$$

3. The first order stationary conditions. In this section, we study the first order stationary conditions of the problem (P).

THEOREM 3.1. *Assume that x^* is a local minimizer of problem (P). Let $J^* = \{1 \leq j \leq n : x_j^* \neq 0\}$, $r^* = |J^*|$. Suppose that the following Robinson condition*

$$\left\{ \left(\begin{array}{c} d_{J^*} - s \\ -d_{J^*} - t \\ d_{\bar{J}^*} \end{array} \right) \middle| \; d \in T_X(x^*), s_{A(x^*)} \leq 0, t_{B(x^*)} \leq 0 \right\} = R^{n+r^*} \tag{3.1}$$

holds, where

$$A(x^*) = \{i \in J^* : x_i^* = b_i\}, \quad B(x^*) = \{i \in J^* : x_i^* = a_i\}.$$

Then there exists $(\nu^*, \eta^*, \kappa^*) \in R^n \times R^n \times R^n$ together with x^* satisfying

$$\begin{aligned} & 0 \in \nabla f(x^*) + \nu^* - \eta^* + \kappa^* + N_X(x^*), \\ & \nu_i^* \geq 0, \quad \eta_i^* \geq 0, \quad \nu_i^*(x_i^* - b_i) = 0, \quad \eta_i^*(-a_i + x_i^*) = 0, \quad i = 1, \ldots, n, \\ & \kappa_j^* = 0, \quad j \in J^*. \end{aligned} \tag{3.2}$$

Proof. The problem (P) can be reformulated to the following problem:

$$\begin{array}{lll} \text{(P)} & \min & f(x) \\ & \text{s.t.} & x \in X, \\ & & a_i \leq x_i \leq b_i, \quad i \in supp(x), \\ & & x_i = 0, \quad i \in \overline{supp}(x), \\ & & |supp(x)| \leq K, \end{array}$$

where $supp(x) = \{1 \leq i \leq n \mid x_i \neq 0\}$, and $\overline{supp}(x)$ is the complement of $supp(x)$ in $\{1, 2, \ldots, n\}$. By the assumption that x^* is a local minimizer of the problem (P), one can observe that x^* is also a local minimizer of the following problem:

$$\begin{array}{lll} (\overline{\text{P}}) & \min & f(x) \\ & \text{s.t.} & x \in X, \\ & & a_i \leq x_i \leq b_i, \quad i \in J^*, \\ & & x_j = 0, \quad j \in \bar{J}^*. \end{array}$$

We observe that the problem $(\bar{P})$ can be reformulated equivalently as

$$\begin{array}{lll} (\overline{\text{P1}}) & \min & f(x) \\ & \text{s.t.} & x \in X, \\ & & I_{J^*}^T (x - b) \leq 0, \\ & & I_{J^*}^T (-x + a) \leq 0, \\ & & I_{\bar{J}^*}^T x = 0. \end{array}$$

Using theorem 3.25 of [22], when the Robinson condition (3.1) holds, there exists $(\nu^*, \eta^*, \kappa^*) \in R^n \times R^n \times R^n$ together with x^* satisfying

$$0 \in \nabla f(x^*) + \nu^* - \eta^* + \kappa^* + N_X(x^*), \tag{3.3}$$

$$\nu_i^* \geq 0, \quad \eta_i^* \geq 0, \quad \nu_i^*(x_i^* - b_i) = 0, \quad \eta_i^*(-a_i + x_i^*) = 0, \quad i \in J^*, \tag{3.4}$$

$$\nu_i^* = 0, \quad \eta_i^* = 0, \quad i \in \bar{J}^*, \tag{3.5}$$

$$\kappa_j^* = 0, \quad j \in J^*. \tag{3.6}$$

Note that (3.4) implies that if $x_i^* \neq 0$ and $x_i^* \neq b_i$, then $\nu_i^* = 0$. If $x_i^* \neq 0$ and $y_i^* \neq a_i$, then $\eta_i^* = 0$. The equation (3.5) implies that if $x_i^* = 0$, then $\nu_i^* = 0$ and $\eta_i^* = 0$. Combining (3.4) with (3.5), we obtain

$$\nu_i^* \geq 0, \quad \eta_i^* \geq 0, \quad \nu_i^*(x_i^* - b_i) = 0, \quad \eta_i^*(-a_i + x_i^*) = 0, \quad i = 1, 2, \ldots, n.$$

It implies that the conclusion holds. □

4. SSAL algorithm. In this section, we propose a SSAL method for solving problem (P). Firstly, we derive the augmented Lagrangian decomposition formulation (P_2) of (P). Secondly, we discuss two subproblems obtained from fixing some variables respectively. Thirdly, we propose the method for solving problem (P) which minimizes (P_2) with respect to x, y in an alternating fashion while updating λ in the iteration. Finally, we establish some convergence results for problem (P).

Introducing $y = x$, we can rewrite (P) as

$$\begin{array}{lll} (\mathrm{P}_1) & \min & f(x) \\ & \text{s.t.} & x \in X, \\ & & y = x, \\ & & \|y\|_0 \leq K, \\ & & y_i \in \{0\} \cup [a_i, b_i], \quad i = 1, 2, \ldots, n. \end{array}$$

Then, we define the augmented Lagrangian function for (P_1) as follows:

$$L_\rho(x, y, \lambda) = f(x) + \lambda^T (y - x) + \frac{\rho}{2}\|y - x\|^2$$

where $\lambda \in R^n$ is the Lagrangian multiplier associated with the equality constraint $y = x$ and $\rho > 0$ is a penalty parameter. The resulting augmented Lagrangian decomposition formulation is:

$$\begin{array}{lll} (\mathrm{P}_2) & \min & L_\rho(x, y, \lambda) \\ & \text{s.t.} & x \in X, \\ & & y \in Y, \end{array}$$

where

$$Y = \{y \in R^n | \ \|y\|_0 \leq K, y_i \in \{0\} \cup [a_i, b_i]\}. \tag{4.1}$$

In the following part, we first discuss the subproblem P_x obtained from fixing the variable y, and then we discuss the subproblem P_y for fixing the variable x.

4.1. The convex programming problem of variables x. We notice that for given $\lambda = \bar{\lambda}$, when $y = \bar{y} \in Y$ is fixed, (P_2) becomes:

$$\begin{array}{lll} (\mathrm{P_x}) & \min & L_\rho(x, \bar{y}, \bar{\lambda}) \\ & \text{s.t.} & x \in X. \end{array}$$

Note that X is a convex set and the objective function of (P_x) is a convex function of x. Thus, (P_x) is a convex programming problem of variables x.

4.2. Minimization of separable quadratic function. We consider the subproblem of (P_2) for given $\lambda = \bar{\lambda}$ when $x = \bar{x} \in X$ is fixed, problem (P_2) becomes

$$\begin{array}{lll} (\mathrm{P_y}) & \min & L_\rho(\bar{x}, y, \bar{\lambda}) \\ & \text{s.t.} & y \in Y. \end{array}$$

The object function $L(\bar{x}, y, \bar{\lambda})$ is a quadratic function of variable y, so we can simplify the objective function $L(\bar{x}, y, \bar{\lambda})$, and then transform the subproblem (P_y) to the $2-$norm approximation problem with semi-continuous variables and the cardinality constraint.

$$\begin{aligned} L_\rho(\bar{x}, y, \bar{\lambda}) &= f(\bar{x}) + \bar{\lambda}^T (y - \bar{x}) + \frac{\rho}{2}\|y - \bar{x}\|^2 \\ &= f(\bar{x}) - \bar{\lambda}^T \bar{x} + \frac{\rho}{2}\|\bar{x}\|^2 + \frac{\rho}{2}[\|y\|^2 - 2(\bar{x} - \frac{1}{\rho}\bar{\lambda})^T y]. \end{aligned}$$

Let $w = \bar{x} - \frac{1}{\rho}\bar{\lambda}$, the object function $L(\bar{x}, y, \bar{\lambda})$ can be rewritten as:

$$
\begin{aligned}
L_\rho(\bar{x}, y, \bar{\lambda}) &= f(\bar{x}) - \bar{\lambda}^T\bar{x} + \frac{\rho}{2}\|\bar{x}\|^2 + \frac{\rho}{2}(\|y\|^2 - 2w^Ty) \\
&= f(\bar{x}) - \bar{\lambda}^T\bar{x} + \frac{\rho}{2}\|\bar{x}\|^2 - \frac{\rho}{2}\|w\|^2 + \frac{\rho}{2}\|y - w\|^2 \qquad (4.2)\\
&= f(\bar{x}) - \frac{1}{2\rho}\|\bar{\lambda}\|^2 + \frac{\rho}{2}\|y - w\|^2
\end{aligned}
$$

For given $\bar{\lambda}$, when $x = \bar{x} \in X$ is fixed, $\frac{1}{2\rho}\|\bar{\lambda}\|^2$, $w = \bar{x} - \frac{1}{\rho}\bar{\lambda}$ are constants. From (4.2), the minimization of the problem (P_y) is equivalent to the following problem :

$$
\begin{aligned}
(\mathrm{P}_3) \qquad \min \quad & \sum_{i=1}^{n}(y_i - w_i)^2 \\
\text{s.t.} \quad & y \in Y.
\end{aligned}
$$

Introducing a variable $z_i \in \{0, 1\}$ to indicate the zero or nonzero status of each decision variable y_i, $i = 1, \dots n$. Problem (P_3) can be reformulated to the following mixed integer program (MIP), which can be solved by CPLEX 12.6.

$$
\begin{aligned}
(\mathrm{MIP}) \qquad \min \quad & \sum_{i=1}^{n}(y_i - w_i)^2 \\
\text{s.t.} \quad & e^Tz \le K, \quad z_ia_i \le y_i \le z_ib_i, \\
& z_i \in \{0, 1\}, \quad i = 1, 2, \dots, n.
\end{aligned}
$$

Let $\bar{y}$ be the optimal solution of the problem (P_y), then there exists $\bar{z}$ together with $\bar{y}$ such that $(\bar{y}, \bar{z})$ is the optimal solution of the problem (P_3). Because the CPLEX 12.6 ignores the feature of this problem (MIP), it is slow, especially for large-scale problems. So we propose new method to obtain the optimal solution $(\bar{y}, \bar{z})$ of (MIP).

Note that the value of variable z_i only has two situations: either $z_i = 0$ or $z_i = 1$. Moreover $z_i = 0$ indicates $y_i = 0$ and $z_i = 1$ indicates $a_i \le y_i \le b_i$, $i = 1, 2, \dots, n$. We can get the optimal solution $(\bar{y}, \bar{z})$ of problem (MIP) successively.

Firstly, taking into account that either $z_i = 1$ or $z_i = 0$, we introduce auxiliary variables q_i and r_i as follows:

$$
q_i = \min_{z_i=0}(y_i - w_i)^2 = \min_{y_i=0}(y_i - w_i)^2 = w_i^2, \text{ where } \bar{y}_i = 0,
$$

$$
\begin{aligned}
r_i &= \min_{z_i=1}(y_i - w_i)^2 = \min_{a_i \le y_i \le b_i}(y_i - w_i)^2 \\
&= \begin{cases} (a_i - w_i)^2 & \text{if} \quad w_i < a_i, \text{ where } \bar{y}_i = a_i, \\ 0 & \text{if} \quad a_i \le w_i \le b_i, \text{ where } \bar{y}_i = w_i, \\ (b_i - w_i)^2, & \text{if} \quad w_i > b_i, \text{ where } \bar{y}_i = b_i. \end{cases}
\end{aligned}
$$

Secondly, we transform the problem (MIP) to the following problem, which is a optimization problem over z :

$$\begin{aligned}
&\min\{\sum_{i=1}^{n}(y_i-w_i)^2 \; |e^Tz\le K, z\in\{0,1\}^n, z_ia_i\le y_i\le z_ib_i, i=1,2,\ldots,n\}\\
&=\min\{\sum_{i=1}^{n} r_iz_i+(1-z_i)q_i \mid e^Tz\le K, z\in\{0,1\}^n\}\\
&=\min\{\sum_{i=1}^{n} q_i+(r_i-q_i)z_i \mid e^Tz\le K, z\in\{0,1\}^n\}\\
&=\sum_{i=1}^{n} w_i^2+\min\{\sum_{i=1}^{n}(r_i-w_i^2)z_i \mid e^Tz\le K, z\in\{0,1\}^n\}
\end{aligned} \tag{4.3}$$

From (4.3), the optimal solution of problem (MIP) is obtained by solving

$$\text{(ILP)} \quad \min\{\sum_{i=1}^{n}(r_i-w_i^2)z_i \mid e^Tz\le K, z\in\{0,1\}^n\}.$$

If $K=n$, the cardinality constraint is void, hence we are left with the simple problem

$$\min\{\sum_{i=1}^{n}(r_i-w_i^2)z_i \; |z\in\{0,1\}^n\},$$

whose optimal solution is given by:

$$\bar{z}_i=\begin{cases}1, & \text{if } r_i-w_i^2\le 0,\\ 0, & \text{otherwise.}\end{cases}$$

In order to deal with the general situation (i.e., if $K<n$) we define $v=r-w^2$. Without loss of generality we assume that v is sorted in ascending order: $v_1\le v_2\le\cdots\le v_n$. Then we have the following theorem.

Theorem 4.1. *The optimal solution $\bar{z}$ of problem (ILP) is given by*

$$\bar{z}_i=\begin{cases}1, & \text{if } i\le K \text{ and } v_i\le 0,\\ 0, & \text{otherwise.}\end{cases}$$

Proof. It is obvious that the optimal value is either negative or zero. If the number of the negative entries of v is less than K, the optimal value is the sum of all negative entries of v, so the optimal solution is:

$$\bar{z}_i=\begin{cases}1, & \text{if } v_i\le 0,\\ 0, & \text{otherwise.}\end{cases}$$

If the number of the negative entries of v is more than K, the optimal value is the sum of the k least entries of v, so the optimal solution becomes.

$$\bar{z}_i=\begin{cases}1, & \text{if } i\le K,\\ 0, & \text{otherwise.}\end{cases}$$

□

After Theorem 4.1, the explicit solution of (P_y) can be given as in the next theorem.

THEOREM 4.2. *For given $\lambda = \bar{\lambda}$ and $x = \bar{x} \in X$, the optimal solution $\bar{y}$ of the problem (P_y) is given by*

$$\bar{y}_i = \bar{z}_i \cdot \min\{b_i, \max\{w_i,\ a_i\}\}, \quad for \quad i = 1, 2, \dots n,$$

where $\bar{z}$ can be obtained from Theorem 4.1.

Yet we are ready to describe our algorithm, named Splitting and Successively solving Augmented Lagrangian method or SSAL method, as in Algorithm 1.

```
Algorithm 1: SSAL method for solving the problem (P)
Input: Choose tolerance parameter $\epsilon \geq 0$, multiplier vector $\lambda^0$, penalty
       parameter $\rho > 0$ and the step-size $\omega$. Set $y^0 \in Y$ and the iteration
       counter $k = 0$.
Output: An approximate optimal solution $(x, z)$ of problem $(P)$.
Solve the subproblem problem (P$_x$) with $\bar{y} = y^0$ to obtain an optimal solution
$x^0$
while $\|x^k - y^k\|_2^2 > \epsilon$ do
    update $z^{k+1}$ and $y^{k+1} = \text{argmin}_{y \in Y} L_\rho(x^k, y, \lambda^k)$, according to
    Theorem 4.1 and Theorem 4.2
    update $x^{k+1} = \text{argmin}_{x \in X} L_\rho(x, y^{k+1}, \lambda^k)$
    update the multipliers by

    (4.4)        $\lambda_i^{k+1} = \lambda_i^k + \omega\rho[y_i^{k+1} - x_i^{k+1}], \quad i = 1, 2, \dots, n$

    increase $k$ by one and continue
return $x = x^k$ and $z = z^k$
```

4.3. Convergence analysis. In this subsection, we show that any accumulation point of the sequence generated by SSAL method satisfies the first order stationary conditions.

THEOREM 4.3. *Let $(\hat{x}, \hat{y}, \hat{\lambda})$ be any accumulation point of (x^k, y^k, λ^k) generated by algorithm 1, $J = \{1 \leq j \leq n : \hat{x}_j \neq 0\}$ and $r = |J|$. Assuming that $\{\lambda^k\}$ is bounded, we have*

$$\lim_{k \to \infty} [(x^{k+1}, y^{k+1}, \lambda^{k+1}) - (x^k, y^k, \lambda^k)] = 0 \tag{4.5}$$

and the Robinson condition (3.1) *holds at $\hat{x}$ for such J . Then, $\hat{x}$ satisfies the first order stationary conditions* (3.2).

The assumption on (4.5) and the boundedness of the multiplier vectors in Theorem 4.3 are the standard condition in the convergence analysis of the augmented Lagrangian methods for nonconvex optimization problems (see [19]). Similar conditions have been used in the convergence analysis of alternating direction methods of multipliers for nonconvex optimization problems (see [23] and [1]).

Before proceeding to the proof of Theorem 4.3, we prove two lemmas.

Lemma 4.4. *Let $J^{k+1} = \{1 \leq j \leq n : y_j^{k+1} \neq 0\}$ and $|J^{k+1}| = r^{k+1}$, then there exists $(\nu^k, \eta^k, \kappa^k) \in R^n \times R^n \times R^n$ together with the x^k, x^{k+1}, y^{k+1} and penalty parameter ρ satisfying:*

$$0 \in \nabla f(x^{k+1}) + \nu^k - \eta^k + \kappa^k + \rho(x^{k+1} - x^k) + N_X(x^{k+1}), \tag{4.6}$$

$$\nu_i^k \geq 0, \eta_i^k \geq 0, \quad \nu_i^k(y_i^{k+1} - b_i) = 0, \eta_i^k(y_i^{k+1} - a_i) = 0, \quad i = 1, 2, \ldots, n, \tag{4.7}$$

$$\kappa_j^k = 0, \quad j \in J^{k+1}. \tag{4.8}$$

Proof. Using (4.3) and the same augment in Theorem 3.1, we have y^{k+1} is also a local minimizer of the following problem:

$$\begin{aligned} (\mathrm{P}_4) \quad \min \quad & \|y - x^k + \frac{1}{\rho}\lambda^k\|^2 \\ \text{s.t.} \quad & I_{J^{k+1}}^T(y - b) \leq 0 \\ & I_{J^{k+1}}^T(-y + a) \leq 0 \\ & I_{\bar{J}^{k+1}}^T y = 0. \end{aligned}$$

Now, this problem is a easy to solve convex optimization problem. And there exists $(\nu^k, \eta^k, \kappa^k)$ together with y^{k+1} satisfying:

$$0 \in \rho(y^{k+1} - x^k + \frac{1}{\rho}\lambda^k) + \nu^k - \eta^k + \kappa^k, \tag{4.9}$$

$$\nu_i^k \geq 0, \eta_i^k \geq 0, \quad \nu_i^k(y_i^{k+1} - b_i) = 0, \quad \eta_i^k(y_i^{k+1} - a_i) = 0, i = 1, \ldots, n, \tag{4.10}$$

$$\kappa_j^k = 0, \quad j \in J^k \tag{4.11}$$

Using the definition of x^{k+1} in algorithm 1 and the Theorem 3.34 of [22], we get the conclusion that x^{k+1} satisfies the following relationship:

$$0 \in \nabla f(x^{k+1}) - \lambda^k - \rho(y^{k+1} - x^{k+1}) + N_X(x^{k+1}). \tag{4.12}$$

Combining (4.9) and (4.12), we obtain

$$\begin{aligned} 0 \in & \rho(y^{k+1} - x^k + \frac{1}{\rho}\lambda^k) + \nu^k - \eta^k + \kappa^k \\ & + \nabla f(x^{k+1}) - \lambda^k - \rho(y^{k+1} - x^{k+1}) + N_X(x^{k+1}), \end{aligned}$$

which implies that

$$0 \in \nabla f(x^{k+1}) + \nu^k - \eta^k + \kappa^k + \rho(x^{k+1} - x^k) + N_X(x^{k+1}).$$

Combining this observation and (4.10), (4.11) , we see that the conclusion holds. □

The proof below uses that $a_i \neq 0$, $b_i \neq 0$. In the compressed sensing problem, we can solving this issue by taking $a_i > 0$, but very small. In practice, this works.

Lemma 4.5. *There exists a subsequence $\hat{K}$ such that $\{(\nu^k, \eta^k, \kappa^k)\}_{k \in \hat{K}}$ is bounded.*

Proof. Since $(\hat{x}, \hat{y}, \hat{\lambda})$ is an accumulation point of (x^k, y^k, λ^k), there exists a subsequence $\{(x^k, y^k, \lambda^k)\}_{k \in \hat{K}} \to (\hat{x}, \hat{y}, \hat{\lambda})$. It follows from (4.4) that $y_i^{k+1} - x_i^{k+1} = (\lambda_i^{k+1} - \lambda_i^k)/\omega\rho$. Taking into account the assumption (4.5), we have

$$\hat{x} = \hat{y}. \tag{4.13}$$

Suppose for contradiction that it is unbounded. By passing to a subsequence if necessary, we can assume that $\|(\nu^k, \eta^k, \kappa^k)\| \to \infty$. Let

$$(\bar{\nu}^k, \bar{\eta}^k, \bar{\kappa}^k) = (\nu^k, \eta^k, \kappa^k)/\|(\nu^k, \eta^k, \kappa^k)\|,$$

then $\|(\bar{\nu}^k, \bar{\eta}^k, \bar{\kappa}^k)\| = 1$. There exists a convergent subsequence $K \subseteq \hat{K}$, such that

$$(\bar{\nu}^k, \bar{\eta}^k, \bar{\kappa}^k) \to (\bar{\nu}, \bar{\eta}, \bar{\kappa}), \quad as \quad k \in K \to \infty, \quad K \subseteq \hat{K},$$

Clearly, $|(\bar{\nu}, \bar{\eta}, \bar{\kappa})\| = 1$, $\bar{\nu} \geq 0$ and $\bar{\eta} \geq 0$. Divide both sides by $\|(\nu^k, \eta^k, \kappa^k)\|$ in (4.6) and take limits as $k \in K \to \infty$. Noticing $f(x)$ is a continuously differentiable convex function and using the the semi continuity of $N_X(\cdot)$ (Lemma 2.42 of [22]), we obtain

$$-\bar{\nu} + \bar{\eta} - \bar{\kappa} \in N_X(\hat{x}) \tag{4.14}$$

Dividing both sides by $\|(\nu^k, \eta^k, \kappa^k)\|$ in (4.7) and (4.8), we have

$$\bar{\nu}_i^k(y_i^{k+1} - b_i) = 0, \quad \bar{\eta}_i^k(y_i^{k+1} - a_i) = 0, \quad i = 1, 2 \ldots, n, \tag{4.15}$$

$$\bar{\kappa}_j^k = 0, \quad j \in J^{k+1}. \tag{4.16}$$

Taking limits as $k \in K \to \infty$ in (4.15), using (4.13), we have

$$\bar{\nu}_i(\hat{x}_i - b_i) = 0, \quad \bar{\eta}_i(\hat{x}_i - a_i) = 0, \quad i = 1, 2, \ldots, n. \tag{4.17}$$

By the definition of J, we have $\hat{x}_i = \hat{y}_i = 0$, for $i \in \bar{J}$. Combining this with (4.17), $a_i \neq 0$ and $b_i \neq 0$, we obtain $\bar{\nu}_i = 0$, $\bar{\eta}_i = 0$. Moreover, we have $\|\bar{\nu}\| = \|\bar{\nu}_J\|$, $\|\bar{\eta}\| = \|\bar{\eta}_J\|$, $\bar{\nu} = I_J\bar{\nu}_J$ and $\bar{\eta} = I_J\bar{\eta}_J$.

Now we show that for $l \in J$, $\bar{\kappa}_l = 0$. For $l \in J$, $\hat{y}_l \neq 0$. Note that $y_l^k \to \hat{y}_l$, there exists a sufficiently large $\hat{k}$, such that $y_l^k \neq 0$ for $k > \hat{k}$. Combining this with (4.16), we have $\bar{\kappa}_l^k = 0$ for $l \in J$, $k > \hat{k}$. Taking limits as $k \in K \to \infty$, we obtain $\bar{\kappa}_l = 0$, for $l \in J$. Thus, we have $\|\bar{\kappa}\| = \|\bar{\kappa}_{\bar{J}}\|$ and $\bar{\kappa} = I_{\bar{J}}\bar{\kappa}_{\bar{J}}$.

Since Robinson's condition (3.1) is satisfied at $\hat{x}$, there exist

$$\tilde{d} \in T_X(\hat{x}), \quad \tilde{s} \in R^r, \quad \tilde{t} \in R^r, \tag{4.18}$$

such that

$$\tilde{s}_{A(\hat{x})} \leq 0, \quad \tilde{t}_{B(\hat{x})} \leq 0, \quad I_J^T\tilde{d} - \tilde{s} = -\bar{\nu}_J, \quad -I_J^T\tilde{d} - \tilde{t} = -\bar{\eta}_J, \quad I_{\bar{J}}^T\tilde{d} = -\bar{\kappa}_{\bar{J}}. \tag{4.19}$$

Using (4.14), (4.18), (4.19), $\|\bar{\nu}\| = \|\bar{\nu}_J\|$, $\|\bar{\eta}\| = \|\bar{\eta}_J\|$, $\|\bar{\kappa}\| = \|\bar{\kappa}_{\bar{J}}\|$, $\bar{\nu} = I_J\bar{\nu}_J$, $\bar{\eta} = I_J\bar{\eta}_J$ and $\bar{\kappa} = I_{\bar{J}}\bar{\kappa}_{\bar{J}}$ we have

$$\begin{aligned}
\|\bar{\nu}\|^2 + \|\bar{\eta}\|^2 + \|\bar{\kappa}\|^2 &= \|\bar{\nu}_J\|^2 + \|\bar{\eta}_J\|^2 + \|\bar{\kappa}_{\bar{J}}\|^2 \\
&= -[(-\bar{\nu}_J)^T\bar{\nu}_J + (-\bar{\eta}_J)^T\bar{\eta}_J + (-\bar{\kappa})_{\bar{J}}^T\bar{\kappa}_{\bar{J}}] \\
&= -[(I_J^T\tilde{d} - \tilde{s})^T\bar{\nu}_J + (-I_J^T\tilde{d} - \tilde{t})^T\bar{\eta}_J + (I_{\bar{J}}^T\tilde{d})^T\bar{\kappa}_{\bar{J}}] \\
&= \tilde{d}^T(-I_J\bar{\nu}_J + I_J\bar{\eta}_J - I_{\bar{J}}\bar{\kappa}_{\bar{J}}) + (s^T\bar{\eta}_J + t^T\bar{\nu}_J) \\
&= \tilde{d}^T(-\bar{\nu} + \bar{\eta} - \bar{\kappa}) + (\tilde{s}^T\bar{\nu} + \tilde{t}^T\bar{\eta}) = \tilde{s}^T\bar{\nu} + \tilde{t}^T\bar{\eta}
\end{aligned}$$

In addition, it is following from (4.17) that if $\hat{x}_i \neq b_i$, then $\bar{\nu}_i = 0$. Thus $\tilde{s}_i\bar{\nu}_i = 0$. If $\hat{x}_i = b_i$, then $\tilde{s}_i \leq 0$. combing this $\bar{\nu} \geq 0$, we have $\tilde{s}_i\bar{\nu}_i \leq 0$. Hence, $\tilde{s}^T\bar{\nu} \leq 0$. By a similar argument as above, one can show that $\tilde{t}^T\bar{\eta} \leq 0$.

Using these relations, we have

$$\|\bar{\nu}\|^2 + \|\bar{\eta}\|^2 + \|\bar{\kappa}\|^2 = \tilde{s}^T\bar{\nu} + \tilde{t}^T\bar{\eta} \leq 0.$$

It yields $\|(\bar{\nu}, \bar{\eta}, \bar{\kappa})\| = 0$, which contradicts $\|(\bar{\nu}, \bar{\eta}, \bar{\kappa})\| = 1$. Therefore, the subsequence $\{(\nu^k, \eta^k, \kappa^k)\}_{k\in\hat{K}}$ is bounded. □

Proof of Theorem 4.3

Proof. Taking into account that $\{(\nu^k, \eta^k, \kappa^k)\}_{k\in\hat{K}}$ is bounded, then there exists a subsequence $\bar{K} \subseteq K$ such that $(\nu^k, \eta^k, \kappa^k) \to (\hat{\nu}, \hat{\eta}, \hat{\kappa})$ as $k \in \bar{K} \to \infty$. Taking limits in (4.6),(4.7) and (4.8) as $k \in \bar{K} \to \infty$, using the relation $\hat{x} = \hat{y}$, the assumption (4.5) and the semicontinuity of $N_{X_0}(\cdot)$, we get

$$\begin{aligned} 0 \in \nabla f(\hat{x}) + \hat{\nu} - \hat{\eta} + \hat{\kappa} + N_X(\hat{x}) \\ \hat{\nu}_i \geq 0, \quad \hat{\eta}_i \geq 0, \quad \hat{\nu}_i(\hat{x}_i - b_i) = 0, \quad \hat{\eta}_i(-a_i + \hat{x}_i) = 0, \quad i = 1, 2, \ldots, n, \\ \hat{\kappa}_i = 0, \quad i \in J \end{aligned}$$

Hence, $(\hat{\nu}, \hat{\eta}, \hat{\kappa})$ together with $\hat{x}$ satisfies the first order stationary conditions (3.2).
□

5. Numerical results. In this section, we conduct numerical experiments to test the performance of SSAL proposed in Section 4. We select the portfolio selection problem (P_{NS}) [13] and the compressed sensing problem (NCSB) as examples. We use real market data to construct the test problems (P_{NS}). We also present simulation study of the performance of SSAL for the portfolio selection problem (P_{NS}) [13] and the compressed sensing problem (NCSB). The experiments demonstrate that as far as the quality of approximate solutions is concerned, SSAL outperforms the PD method and is almost as good as the CPLEX and the MIQCQP_1. At the same time, the computational time for SSAL is significantly smaller than the other three methods. The numerical tests are implemented in MATLAB 7.14 and run on a PC (2.10G Hz, 2GB RAM).

5.1. Portfolio selection problem with nonsystematic risk constraint. In this subsection, we verify the efficiency and stability of SSAL method in portfolio selection problem (P_{NS}) of [13]:

$$\begin{aligned} (\mathrm{P_{NS}}) \qquad \min \quad & f(x) = x^T(Q + D)x \\ \text{s.t.} \quad & g(x) = x^T D x \leq \sigma_0, \\ & \alpha^T x \geq \rho_0, \quad e^T x = 1, \\ & \|x\|_0 \leq K, \\ & x_i \in \{0\} \cup [a_i, b_i], \quad i \in \{1, 2, \ldots, n\}, \end{aligned}$$

where Q is a semi-definition matrix and D is a diagonal matrix. Clearly, problem (P_{NS}) is in the form of (P), thus SSAL proposed in Section 4 can be suitably applied to solve (P_{NS}). The $x-$update step involves a quadratically constrained quadratic program (QCQP), which is

$$\begin{aligned} (\mathrm{QCQP}) \qquad \min \quad & x^T\left(Q + D + \frac{\rho}{2}E\right)x - \left(\bar{\lambda} + \rho\bar{y}\right)^T x \\ \text{s.t.} \quad & g(x) = x^T D x \leq \sigma_0, \quad \alpha^T x \geq \rho_0, \\ & e^T x = 1, \quad 0 \leq x \leq b. \end{aligned}$$

We use the CPLEXQCP solver in CPLEX 12.6 [12] with MATLAB interface to solve the (QCQP) problem. The $y-$update step has exactly the same expression as in the Theorem 4.2. It is obviously that the main computational effort of SSAL lies in solving the subproblem (QCQP). In our test, the parameters in SSAL method are set as follows: the step-size $\omega = 0.3$, the initial Lagrangian multiplier $\lambda^0 = 0$ and the initial penalty $\rho = 1$. We set the prescribed weekly return level $\rho_0 = 2 \times 10^{-3}$, the prescribed nonsystematic risk level $\sigma_0 = 1 \times 10^{-3}$, $a_i = 0.01$, and $b_i = 0.3$, $i = 1, 2, \ldots, n$. The cardinality upper bound $K = 10$, and the tolerance parameter is set as $\epsilon = 10^{-4}$.

5.1.1. Real data set. In this subsection, we use the weekly return data of 481 stocks from Standard&Poor's 500 index between 2005 and 2010. We take linear regressions on 10 sector indexes of Standard&Poor's 500 index to construct a sector factor model. The objective function of (P$_{NS}$) is the covariance obtained by the sector factor model while the quadratic constraint controls the nonsystematic risk $x^T Dx$ in the sector factor model. We generate 5 instances for each test problem with the same size (n=50,100,150). For each instance, we choose n stocks from the 481 stocks randomly.

In order to test the accuracy of solution obtained by SSAL method, we compare SSAL with the CPLEXMIQCP solver in CPLEX 12.6 [12]. Introducing a variable $z_i \in \{0, 1\}$ to indicate the zero or nonzero status of each decision variable x_i, $i = 1, 2, \ldots, n$, problem (P_{NS}) can be reformulated as follows:

$$\begin{array}{lll} (\mathrm{MIP}_0) & \min & x^T(Q+D)x \\ & \text{s.t.} & x^T Dx \le \sigma_0, \\ & & \alpha^T x \ge \rho_0, \quad e^T x = 1, \\ & & e^T z \le K, \quad z_i a_i \le x_i \le z_i b_i, \\ & & z_i \in \{0,1\}, \quad i = 1, 2, \ldots, n. \end{array}$$

The experiments are conducted by using the default setting of CPLEX 12.6 with the maximum CPU time set equal to 3600 s.

To illustrate the performance of SSAL for test problems (P$_{NS}$) specifically, we present three comparison results in terms of the computational time in seconds, the number of iteration and the objective value explored by SSAL method and CPLEX 12.6 in Table 1. We also record the relative difference (rel.dif) in Table 1, where rel.dif:=$(fval(x_{SSAL}) - fval\,(x_{CPLEX}))\,/fval(x_{CPLEX})$.

The numerical results in Table 5.1 show that the cost of computational efforts of SSAL is significantly less than the well-known solver CPLEX 12.6 for solving (P$_{NS}$) on a real world data. For instance, the example listed the last row in the Table 1 shows that CPU time of SSAL is almost 200 times faster than that of the CPLEX. Obviously, we conclude that the SSAL method outperforms the solver CPLEX 12.6. Note that the optimal objective value of the problem (P$_{NS}$) solved by SSAL is almost same to that of the CPLEX. The largest absolute value of the relative difference (rel.dif) is less 4%.

Then we continue to compare SSAL method with the methods provided by Cui

n	SSAL			CPLEX		
	fval×10^{-4}	cputime	iter	fval×10^{-4}	cputime	rel.dif×10^{-2}
50	2.3503	0.3746	7	2.2981	0.9093	2.2725
50	2.7138	0.1697	5	2.6584	1.4307	2.0845
50	3.7464	0.2343	5	3.6557	2.0024	2.4788
50	3.5357	0.6295	16	3.5446	2.7720	-0.0507
50	2.1145	0.2257	7	2.0819	1.3988	1.5695
100	2.3550	0.7521	8	2.3294	9.8475	1.0969
100	1.4998	0.7698	10	1.5232	15.8758	-1.5421
100	4.3390	0.9155	10	4.3077	6.5231	0.7267
100	2.0593	1.0874	14	2.0380	46.6908	1.0468
100	1.6443	1.3415	15	1.6683	84.1186	-1.4396
150	1.4061	2.1155	11	1.3803	168.8789	1.8750
150	2.8602	1.4629	8	2.8579	34.4529	0.0799
150	1.8396	2.1173	13	1.8023	312.4525	2.0674
150	1.4054	3.7521	16	1.3631	497.5107	3.1043
150	1.2593	3.9929	18	1.2492	732.9191	0.8100

Table 5.1: SSAL versus CPELX 12.6 on the (P_{NS}) problem

et al in [13], where a new model defined as MIQCQP$_1$ as follows:

$$
\begin{aligned}
(\mathrm{MIQCQP}_1) \qquad \min \qquad & f(x,\delta) = x^T(Q+D)x + \sum_{i-1}^{n} d_i\delta_i \\
\text{s.t.} \quad & g(x) = \sum_{i=1}^{n} d_i\delta_i \le \sigma_0, \quad \alpha^T x \ge \rho_0, \\
& e^T x = 1, \quad \|x\|_0 \le K, \quad x_i^2 \le \delta_i z_i, \\
& x_i \in \{0\} \cup [a_i, b_i], \quad i \in \{1,2,\dots,n\}.
\end{aligned}
$$

Problem (MIQCQP$_1$) is also solved by the mixed integer QCP solver in CPLEX 12.6 with MATLAB interface using continuous relaxation for generating lower bounds. Cui et al claimed that their MIQCQP$_1$ is more efficient than the CPLEX.

Now we generate the above test problems with n ranged from 100 to 400. For each n, we generate 20 instances by selecting n stocks from the 481 stocks randomly.

We illustrate the numerical results by mean of Figure 5.1, comparing the average computational time of SSAl with that of (MIQCQP$_1$). Figures 5.1 shows that SSAL also promises (MIQCQP$_1$). For example, our method is 20 times faster than that of (MIQCQP$_1$), under the best situation.

5.1.2. Simulation data set. To further demonstrate the effective of performance of SSAL for large-scale problems, in this subsection, we simulate the test examples with the large size from 500 to 1000. The simulation data is generated in the same fashion as in [26]. The parameters in the model (P_{NS}) are generated randomly from the uniform distributions: $\alpha_i \in [0, 0.03]$, $\beta_{ij} \in [0.3, 2]/m$, $i = 1, 2, \dots n$, $j = 1, 2, \dots m$; σ_{ij}, the covariance of factors i and j is calculated with sampled series from $[0, 0.4]$ randomly, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, m$; $\sigma_{\epsilon_i} \in [0, 0.002]$, $i = 1, 2, \dots n$.

Our goal is to explore the tendency of the computational time as n increasing. 10

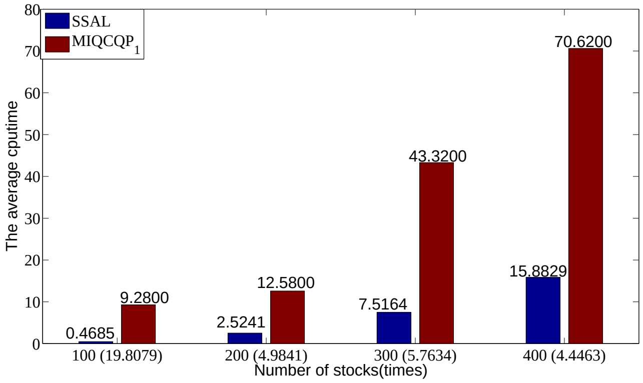

Fig. 5.1: Average CPU time (in seconds) for problems (P$_{NS}$) of SSAL and ($MIQCQP_1$)

n	Computating time (s)			Iteration		
	minimum	maximum	average	minimum	maximum	average
500	21.0029	48.8297	29.82437	8	15	12
600	36.9216	57.6063	47.9425	8	11	10
700	68.4558	103.8730	85.6297	9	12	11
800	126.9821	160.2004	140.9647	8	10	9
900	173.8905	253.4003	212.5805	8	10	10
1000	249.5224	493.6785	355.7434	8	14	11

Table 5.2: The variability for the large-scale problems

instances are generated randomly and recorded the three indexes: minimum, maximum and average numbers of iterations and computational time, respectively. It is importance to note that CPLEX 12.6 is not able to solve the most of instances within 3600s.

It is clear from Table 5.2 that SSAL is robust and efficient as the number of assets increases. In addition, the number of iteration less depends on the problem size.

5.2. Compressed sensing and subset selection. In this subsection, we apply SSAL method proposed in Section 4 to solve the problem (NCSB). We generate data as follows: the data matrix $A \in R^{p\times n}$ with orthonormal rows, each entry from a standard Gaussian distribution $a_{ij} \sim \mathcal{N}(0,1)$. The original signal f is generated with K nonzero entries, each sampled is the absolute value of $\hat{x}$, where $\hat{x}$ from an $\mathcal{N}(0,1)$ distribution. An observation vector $b = Af + r \in R^p$, where r is Gaussian noise of variance $\sigma^2 = 0.01$.

Firstly, we illustrate the behavior of SSAL visually. The x-update step involves an unconstrained quadratic optimization problem P_x. The $y-$update step has exactly the same expression as in the Theorem 4.2, where the semicontinuous bound $a_i = 1\times 10^{-5}$, $i = 1, 2, \ldots, n$. We generate one instance with $(p, n) = (1024, 2048)$, $K = 500$ randomly. The parameters in SSAL are set as follows: the initial Lagrangian multiplier $\lambda^0 = 0$, the tolerance parameter is set as $\epsilon = 10^{-4}$, $u = 10$, the step-size $\omega = 1$ and the initial penalty $\rho = 1$. We use notations "Original" and "SSAL" as the original signal,

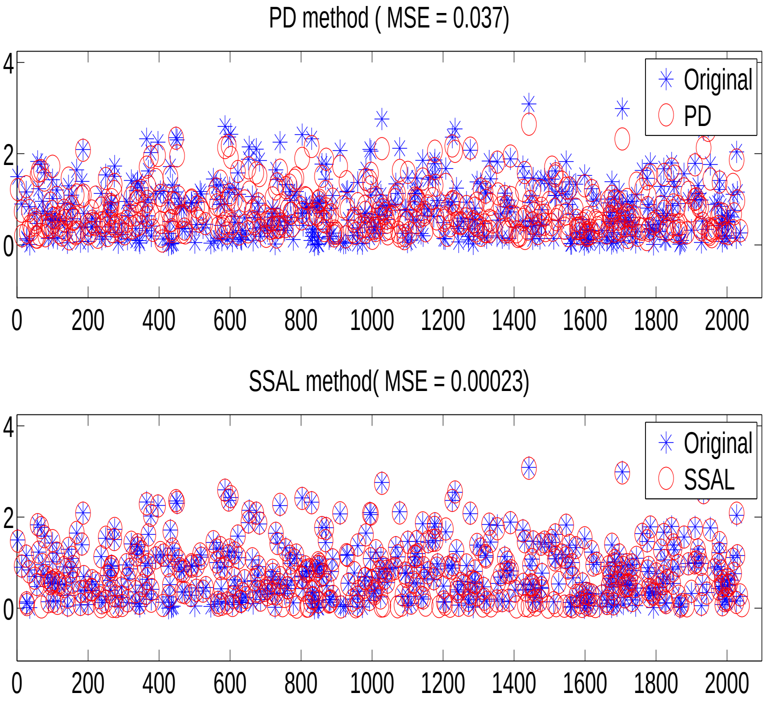

Fig. 5.2: Compare the performance of SSAL method and the PD method to the problem (NCSB)

the signal recovered by the SSAL respectively. From the bottom graph of Figure 5.2, it is clear that our method almost recover the original signal exactly.

We also compare the performance of SSAL with the PD method of [18] for this simple test. For the PD method, we set the tolerance parameter $eps = 10^{-6}$ and the initial point as 0. To evaluate the quality of these sparse approximate solutions, we adopt a similar criterion as described in [14]. The associated squared error is defined as:

$$MSE := \frac{1}{n}\|f - \hat{f}\|_2^2,$$

where $\hat{f}$ is an estimate of the original signal f.

Comparing the top graph to the bottom one in Figure 5.2, we observe that SSAL is better recover the signal than the PD method, and the MSE of the approximate solutions obtained by SSAL are much lower than that of the PD method.

In order to illustrate that SSAL is the winner, we compare SSAL with the PD method [18] with two different sizes. We construct the testing data and then set of all the parameters values in the same way as the previous test. For each size, we apply SSAL and the PD method to solve problem (NCSB) with four different values of K. For each such K, we generate the data set consisting of 50 instances randomly. The first 50 problems with $K = 50$, then $K = 100$, $K = 150$ and $K = 200$. Figure 5.3 and Figure 5.4 show the 10-logarithm of the ratio t_{PD}/t_{SSAL} and the 10-logarithm

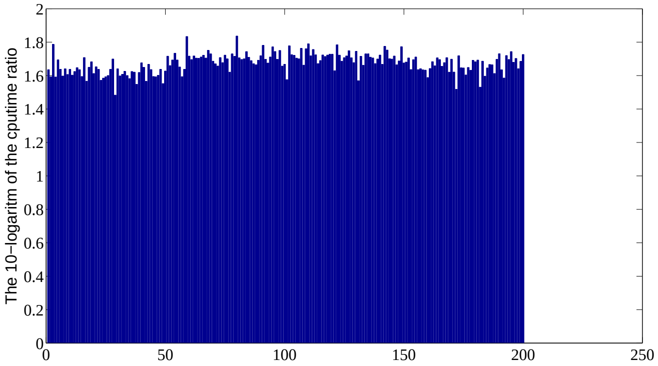

Fig. 5.3: The 10-logarithm of the ratio $\frac{t_{PD}}{t_{SSAL}}$ for $(p, n) = (512, 1024)$

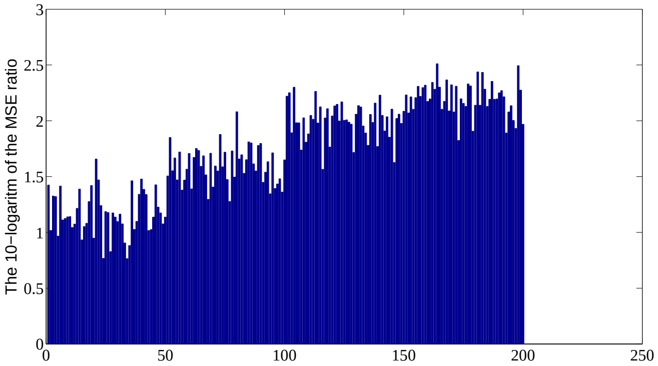

Fig. 5.4: The 10-logarithm of the MSE $\frac{MSE_{PD}}{MSE_{SSAL}}$ for $(p, n) = (512, 1024)$

of the ratio MSE_{PD}/MSE_{SSAL} for $(p, n) = (512, 1024)$, respectively. Figure 5.5 and Figure 5.6 show the 10-logarithm of the ratio t_{PD}/t_{SSAL} and the 10-logarithm of the ratio MSE_{PD}/MSE_{SSAL} for $(p, n) = (1024, 2048)$, respectively.

From Figure 5.3 and Figure 5.5, we observe that SSAL is substantially faster than the PD method for every case. Moreover, our method is 40 times faster than the PD method on average. For the best situation, our method is more than 100 times faster than the PD method. In addition, SSAL outperforms the PD method in terms of solution quality since it archives much smaller MSE values. Thus, SSAL much better recovers the original signal than the PD method.

Empirical study demonstrates that SSAL is a powerful method to solve problem (P) for the following reasons: (1) SSAL can find the good enough approximate solution of problem (P). For example, the optimal objective value obtained by the SSAL and the CPLEX are almost the same for the problem (P_{NS}). Furthermore, the quality of the approximate solution of problem (NCSB) obtained by SSAL is generally better than that of the PD method. (2) SSAL is a very fast method. For instance, SSAL is 200 times faster than the CPLEX method and 5 times faster than the (MIQCQP_1) method of [13] for the problem (P_{NS}). For the (NCSB) problem, SSAL is 40 times

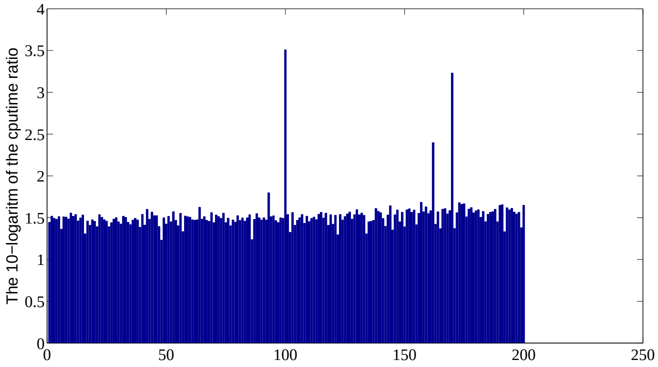

Fig. 5.5: The 10-logarithm of the ratio $\frac{t_{PD}}{t_{SSAL}}$ for $(p, n) = (1024, 2048)$

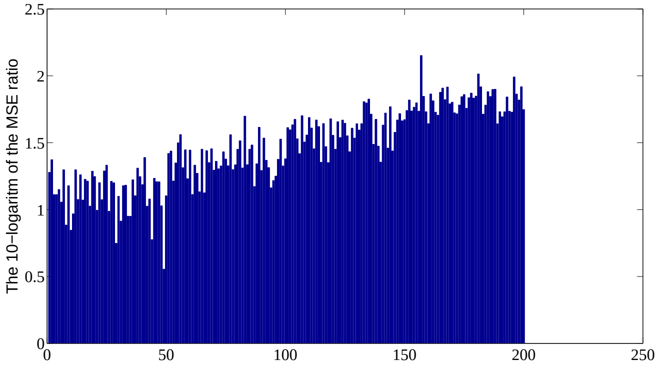

Fig. 5.6: The 10-logarithm of the MSE $\frac{MSE_{PD}}{MSE_{SSAL}}$ for $(p, n) = (1024, 2048)$

faster than the PD method of [18].

6. Conclusion. We have proposed a splitting and successively solving augmented Lagrangian (SSAL) method. Subproblems have been decomposed by fixing certain variables and solving them at each iteration alternatively. The convergence of SSAL has been proved, under some suitable assumptions. Real world data and simulation study show that SSAL outperforms the CPLEX 12.6, the (MIQCQP$_1$) reformulation [13] and the PD method [18], while enjoying a similar cardinality of decision variable. In particularly, SSAL is powerful when the size of the problem increase largely.

REFERENCES

[1] Xiaodi Bai, Jie Sun, and Xiaoling Sun. An alternating direction method for chance-constrained optimization problems with discrete distributions. *Submitted to Comput. Oper. Res. ,under review*, 2013.

[2] Yanqin Bai and Kaiji Shen. Alternating direction method of multipliers for $\ell_1 - \ell_2$ regularized

logistic regression model. *J. Oper. Res. Soc. China*, 3:DOI 10.1007/s40305–015–0090–2, 2015.

[3] Yanqin Bai, Yanjun Shen, and Kaiji Shen. Consensus proximal support vector machine for classification. *J. Oper. Res. Soc. China*, 2:57–74, 2014.

[4] Dimitris Bertsimas and Romy Shioda. Algorithm for cardinality-constrained quadratic optimization. *Comput. Optim. Appl.*, 43(1):1–22, 2009.

[5] Daniel Bienstock. Computational study of a family of mixed-integer quadratic programming problems. *Math. Program.*, 74(2, Ser. A):121–140, 1996.

[6] Thomas Blumensath and MikeE. Davies. Iterative thresholding for sparse approximations. *J. Fourier Anal. Appl.*, 14(5-6):629–654, 2008.

[7] Pierre Bonami and Lejeune Miguel. An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Oper. Res.*, 57(3):650–670, 2009.

[8] Oleg Burdakov, Christian Kanzow, and Alexandra Schwartz. Mathematical programs with cardinality constraints : Reformulation by complementarity-type constraints and a regularization method. Technical report, Linkӧping University, Optimization, 2014.

[9] Oleg Burdakov, Christian Kanzow, and Alexandra Schwartz. On a reformulation of mathematical programs with cardinality constraints. *Springer International Publishing*, 95:3–14, 2015.

[10] Tommy Chang, Nigel Meade, John E. Beasley, and Yazid M. Sharaiha. Heuristics for cardinality constrained portfolio optimisation. *Comput. Oper. Res.*, 27(13):1271 – 1302, 2000.

[11] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.

[12] IBM ILOG CPLEX. IBM ILOG CPLEX Optimization Studio V12.6.0, 2013.

[13] Xueting Cui, Xiaojin Zheng, Shushang Zhu, and Xiaoling Sun. Convex relaxations and MIQCQP reformulations for a class of cardinality-constrained portfolio selection problems. *J. Global Optim.*, 56(4):1409–1423, 2013.

[14] Mario A. T. Figueiredo, Robert D. Nowak, and Stephen J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Signal Process.*, 1(4):586–597, Dec 2007.

[15] Antonio Frangioni and Claudio Gentile. Perspective cuts for a class of convex 0-1 mixed integer programs. *Math. Program.*, 106(2, Ser. A):225–236, 2006.

[16] Jianjun Gao and Duan Li. A polynomial case of the cardinality-constrained quadratic optimization problem. *J. Global Optim.*, 56(4):1441–1455, 2013.

[17] M. Amin Khajehnejad, Alexandros G. Dimakis, Weiyu Xu, and Babak Hassibi. Sparse recovery of nonnegative signals with minimal expansion. *IEEE Trans. Signal Process.*, 59(1):196–208, 2011.

[18] Zhaosong Lu and Yong Zhang. Sparse approximation via penalty decomposition methods. *SIAM J. Optim.*, 23(4):2448–2478, 2013.

[19] H. Z. Luo, X. L. Sun, and D. Li. On the convergence of augmented Lagrangian methods for constrained global optimization. *SIAM J. Optim.*, 18(4):1209–1230, 2007.

[20] P.D. O'Grady and Scott T. Rickard. Compressive sampling of non-negative signals. In *Machine Learning for Signal Processing, 2008. MLSP 2008. IEEE Workshop on*, pages 133–138, Oct 2008.

[21] Manfred Padberg and Giovanni Rinaldi. A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems. *SIAM Rev.*, 33(1):60–100, February 1991.

[22] Andrzej Ruszcunski. *Nonlinear Optimization*. Princeton University Press, 2006.

[23] Y. Shen, Z. Wen, and Y. Zhang. Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization. *Optim. Methods Softw.*, 29(2):239–263, 2014.

[24] Xiaoling Sun, Xiaojin Zheng, and Duan Li. Recent advances in mathematical programming with semi-continuous variables and cardinality constraint. *J. Oper. Res. Soc. China*, 1(1):55–77, 2013.

[25] Xiaojin Zheng, Xiaoling Sun, and Duan Li. Improving the performance of MIQP solvers for quadratic programs with cardinality and minimum threshold constraints: A semidefinite program approach. *INFORMS J. Comput.*, 26(4):690–703, 2014.

[26] Shushang Zhu and Xueting Cui. Factor-risk constrained mean-variance portfolio selection:formulation and global optimization solution approach. *Journal of Risk*, 14(2):51–89, 2011.